Obtaining a standardised CPUE series for toothfish (*Dissostichus eleginoides*) in the Prince Edward Islands EEZ calibrated to incorporate both longline and trotline data over the period 1997-2012

Anabela Brandão and Doug S. Butterworth

Marine Resource Assessment & Management Group (MARAM)
Department of Mathematics and Applied Mathematics
University of Cape Town
Rondebosch, 7701, South Africa

October 2012

Abstract

The previous GLMM standardisation approach for PEI toothfish (Spanish) longline CPUE data is extended to include data for the 2011-2012 seasons and the same approach is applied to trotline CPUE data for the 2008–2012 period. CPUE data from a research program carried out in 2012 in which longline and trotline sets were paired to within three nautical miles and a period of two weeks is analysed to obtain a calibration factor for longlines and trotlines. A model is then fitted to combine the two individual standardised CPUE series and the calibration factor to obtain a "calibrated" longline CPUE series (incorporating both longline and trotline information) and an estimate of the calibration factor. This indicates a drop of about 44% in standardised CPUE in 2012 compared to the immediately preceding years; however it must be noted that the data in 2012 is available only until August.

Introduction

The General Linear Mixed Model (GLMM) of Brandão and Butterworth (2011) has been applied to standardise the longline (Spanish) CPUE data for toothfish in the Prince Edward Islands EEZ for which data are now available until August 2012. The same form of GLMM has also been applied to the trotline CPUE data that are available since 2008.

A GLM analysis has been performed on paired longline and trotline CPUE data obtained from a research program carried out in 2012 to obtain a calibration factor between the two types of gear. Results from these three analyses are then modelled jointly to obtain a calibrated longline CPUE series for the 1997 to 2012 period.

The Data

Longline commercial catch data (as kg green weight), and effort data (as total number of hooks set) are available for the period 1997 to August 2012 and a total of 7 745 sets are available for analyses (Table 1a). Trotline CPUE data are available for the 2008 to August 2012 period. The effort for a trotline is defined as

$$\left(\frac{\text{Length of line}}{\text{Spacing of droppers}}\right)\times \text{Number of clusters per dropper.}$$

A total of 753 trotline sets (Table 1b) are available for analyses.

In 2012 a research program was carried out in which longline and trotline sets had to meet certain criteria. In main, they had to be set within three nautical miles and within a period of two weeks of each other. A total of 47 pairs of such data are available for analyses.

It has come to light that there are some questions about the accuracy of the commercial CPUE that is available from different sources (such as the data used in previous CPUE analyses, the CCAMLR database and the original C2 forms and observer forms). In an initial comparison of the data from different sources, one difference that has been corrected in the data used in the present analyses is that previously some sets were accorded a zero catch but in the CCAMLR database they are recorded “NA” sets indicating that the set has no catch for some reason presumably unrelated to local abundance of toothfish. All these sets have now been omitted from the analyses. A full data verification exercise has not been possible due to a lack of time before analyses had to be completed to provide a basis for management advice.

Methods

GLMM model to standardise CPUE data

Brandão and Butterworth (2011) proposed that a General Linear Mixed Model (GLMM) be used in the standardisation of the toothfish CPUE in which all interaction terms are considered as random effects because of the low number of longline sets (Table 1a) in the most recent years, because otherwise a large number of interaction terms have to be set using interpolation. In this paper GLMMs have been used to standardise the longline as well as the trotline commercial CPUE data.

The GLMM applied to the longline (and to trotline) CPUE data is of the form:

$$\ln\left(CPUE + \delta\right) = X\alpha + Z\beta + \varepsilon \, , \qquad (2)$$

where

CPUE is the longline/trotline catch per unit effort,

δ is a small constant (10% of the average of all CPUE data values = 0.029) added to the toothfish CPUE to allow for the occurrence of zero CPUE values,

α is the unknown vector of fixed effects parameters which includes: $\mu + \kappa_{vessel} + \omega_{year} + \gamma_{month} + \lambda_{area}$, where

- μ is the intercept,
- *vessel* is a factor with 9 levels associated with each of the vessels that have operated in the fishery (to an appreciable extent):

  Aquatic Pioneer
  Arctic Fox
  El Shaddai
  Eldfisk
  Isla Graciosa
  Koryo Maru 11
  Ross Mar
  South Princess
  Suidor One

year — is a factor with 16 levels associated with the years 1997–2012 for longlines or with 5 levels associated with the years 2008–2012 for trotlines,

month — is a factor with 12 levels (January– December),

area — is a factor with 4 levels associated with the four spatially distinct fishing areas:

A: 43–48°S latitude and 32–37°E longitude,

B: 43–45.3°S latitude and 37–40.3°E longitude,

C: 45.3–48°S latitude and 37–40.3°E longitude,

D: 43–48°S latitude and 40.3–43.3°E longitude,

X — is the design matrix for the fixed effects,

β — is the unknown vector of random effects parameters which includes the following interaction terms:

$\eta_{year \times area} + \theta_{year \times month} + \phi_{month \times area}$, where

year×area — is the interaction between year and area (this allows for the possibility of different variation with time for the different areas),

year×month — is the interaction between year and month,

month×area — is the interaction between month and area,

Z — is the design matrix for the random effects, and

ε — is an error term assumed to be normally distributed and independent of the random effects.

It is assumed that both the random effects and the error term have zero mean, i.e. E(β) = E(ε) = 0, so that $E(\ln(CPUE + \delta)) = \mathbf{X}\alpha$. We denote the variance-covariance matrix for the residual errors (ε) by **R** and the variance-covariance matrix for the random effects (β) by **G**. In the analyses of this paper it is assumed that the residual errors as well as the random effects are homoscedastic and are uncorrelated, so that both **R** and **G** are diagonal matrices given by:

$$\mathbf{R} = \sigma_{\varepsilon}^{2}\mathbf{I}$$
$$\mathbf{G} = \sigma_{\beta}^{2}\mathbf{I}$$

where **I** denotes an identity matrix. Thus, in the mixed model, the variance-covariance matrix (**V**) for the response variable is given by:

$$\mathrm{Cov}(\ln(CPUE+\delta)) = \mathbf{V} = \mathbf{ZGZ}^{\mathrm{T}} + \mathbf{R},$$

where $\mathbf{Z}^{\mathrm{T}}$ denotes the transpose of the matrix **Z**.

The estimation of the variance components (**R** and **G**), the fixed effects (α) and the random effects (β) parameters in GLMM requires two steps. First the variance components are estimated. Once estimates of **R** and **G** have been obtained, estimates for the fixed effects parameters (α) can be obtained as well as predictors for the random effects parameters (β). Variance component estimates are obtained by the method of residual maximum likelihood (REML) which produces unbiased estimates for the variance components as it takes the degrees of freedom used in estimating the fixed effects into account.

To provide additional insight GLMM analysis was also performed by introducing an extra "gear" fixed factor to incorporate CPUE data from both longlines and trotlines to obtain an estimate of a "gear" effect. In this instance we are ignoring the pairing of some of the longline and trotline sets in 2012 that were part of a research program for the purposes of getting a calibration factor for longlines and trotlines, so that the information content of these paired sets as regards the calibration factor is underweighted.

GLM to analyse research paired CPUE data from longlines and trotlines

The GLM considered allows for possible differences in "catchability" between the two types of gear (i.e. different multiplicative bias factors g) as well as for varying spatial and temporal distribution of toothfish density. The model is thus given by:

$$\ln\left(CPUE+\delta\right) = \mu + \alpha_g + \beta_{pair} + \varepsilon,$$

where

- *CPUE* is the catch per unit effort for longlines or trotlines, where the effort for the different gears are described earlier in the paper,
- δ is a small constant (10% of the average of the paired CPUE data values = 0.085) added to the toothfish CPUE to allow for the occurrence of zero CPUE values,
- μ is the intercept (which incorporates the longline gear factor),
- g is a factor with 2 levels associated with the type of gear (longline or trotline),
- *pair* is a factor with 47 levels associated with set pairs between the Spanish longline and the trotline gear (capturing the different areas and times that

the experiments took place, for each of which the underlying toothfish density may have been different), and

ε is the error term assumed to be normally distributed.

Since $\alpha_{longline}$ is incorporated in the intercept, the (log-transformed) calibration factor from this analysis, $K^* = \alpha_{trotline}$, with the analysis providing an estimate of K^* and of its associated variance $\sigma^2_{K^*}$.

During the research sets cetacean predation was observed to a much higher extent by the observers on one vessel than on the other vessel. However this information has not been included in the analyses because the information recorded is only whether cetaceans in the vicinity were observed to be feeding on the toothfish or not. This information is recorded only by the observers on the vessels so not every set has this information. Also, cetaceans could be feeding on the toothfish underwater and therefore not be seen to be feeding by the observer.

Model to calibrate the standardised longline CPUE series given the standardised trotline CPUE

The following negative log-likelihood function is minimised to estimate a calibrated longline CPUE series:

$$-\ln L = \frac{1}{2}\ln\left|\mathbf{V}_L\right| + \frac{1}{2}\left(\ln \mathbf{CPUE}^L - \ln \mathbf{CPUE}^{cal}\right)^T \mathbf{V}_L^{-1}\left(\ln \mathbf{CPUE}^L - \ln \mathbf{CPUE}^{cal}\right) +$$
$$\frac{1}{2}\ln\left|\mathbf{V}_T\right| + \frac{1}{2}\left(\ln \mathbf{CPUE}^T - \ln\mathbf{K} - \ln \mathbf{CPUE}^{cal}\right)^T \mathbf{V}_T^{-1}\left(\ln \mathbf{CPUE}^T - \ln\mathbf{K} - \ln \mathbf{CPUE}^{cal}\right) +$$
$$\frac{1}{2\sigma^2_{K^*}}\left(\ln K^* - \ln K\right)^2$$

where

$\mathbf{CPUE}^{L/T}$ is the vector of the predicted longline/trotline CPUE values obtained from fitting the GLMM described earlier to obtain standardised longline/trotline CPUE series,

$\mathbf{V}_{L/T}$ is the variance-covariance matrix of the predicted longline/trotline (log) CPUE series,

CPUE^{cal} is the vector of estimated longline CPUE which incorporates the calibrated trotline data,

K is the vector of the estimated calibration factor between longlines and trotlines (this is defined as a vector for the purposes of conducting vector/matrix calculations but it contains only one value, indicated as *K* in the last part of the equation above),

*K** is the calibration factor obtained from analysing the paired research CPUE data, and

$\sigma^2_{K^*}$ is the variance of the *K** parameter.

It might appear that the data from the paired longline/trotline sets are being used twice in this likelihood. Note however that the GLMM analyses inputs in the first two lines of the RHS take only the trend information in these data into account, whereas their information in regard to the method calibration is taken into account only by the term in the final line.

Results and Discussion

Table 2 and Figure 1 show the relative abundance indices for toothfish provided by the standardised commercial longline and trotline (calibrated to longline) CPUE series for the Prince Edward Islands EEZ as well as the longline CPUE series calibrated by the trotlines. There is a large difference between the 2011 index from the longline GLMM and the calibrated index. It should be noted, however, that there were only two longline sets in 2011 (see Table 1a), so that appropriately the calibrated index is very close to the estimate related to the trotline data for this year. Figure 2 reproduces the CPUE series of Figure 1 individually with their 95% confidence intervals.

Table 3 gives the parameter estimates and their 95% confidence intervals for the three variants of the “gear” factor.

For assessment purposes, the appropriate index of abundance to use is the calibrated series listed in Table 2 and shown in the top plot of Figure 2. A concern is that the point estimate for 2012 is well below the values for the previous two years and the third lowest in the full time series; however the data available for 2012 do extend only to August.

Acknowledgements

Thanks to Deon Durholtz and Rob Leslie of DAFF and Chris Heinecken of CAPFISH, for assistance with queries about the data upon which this paper is based.

Reference

Brandão, A. and Butterworth, D.S. 2011. General Linear Model and General Linear Mixed Model standardisation of the toothfish (*Dissostichus eleginoides*) CPUE in the Prince Edward Islands vicinity over the period 1997–2010. DAFF Branch Fisheries document: FISHERIES/2011/OCT/SWG-DEM/48.

Table 1a. The number of data entries per month and year available for the GLMM analysis to standardise the commercial Spanish longline toothfish CPUE series.

Year	Month												Total
	1	2	3	4	5	6	7	8	9	10	11	12	
1997	34	80	11	11	72	25	12	38			76	113	472
1998	134	173	194	150	14	37	126	93	89	231	64	81	1386
1999	48	34	30	69	183	65	13	194	212	131	198	54	1231
2000	153	200	138	172	149	66	162	142	148	199	95	47	1671
2001		39	55	14	121	155	53	90	5	28	15	9	584
2002	5	47	69	15	11				34	73	65		319
2003	2	35	47		17	90	111		43	183	45		573
2004		15	53		50	128	30		5	55	61	20	417
2005		10	59	3					14	52	43		181
2006		18	43					7	39	30			137
2007		35	52	22	135	65	21	87	10	43	39		509
2008		23	33							21	12		89
2009				2						21	31		54
2010	2	34	37										73
2011				2									2
2012				7	11		24	5					47

Table 1b. The number of data entries per month and year available for the GLMM analysis to standardise the commercial trotline toothfish CPUE series.

Year	Month												Total
	1	2	3	4	5	6	7	8	9	10	11	12	
2008			5							9	45	2	61
2009										24	28		52
2010		5	18	2					12	69	60	5	171
2011	29	2	50	44	30		11	17	73	16	40	15	327
2012				20	38		33	51					142

Table 2. Relative abundance indices for toothfish provided by the standardised commercial CPUE series for the Prince Edward Islands EEZ for the Spanish longline and for the trotline fisheries, and the calibrated longline CPUE series.

Year	GLMM *CPUE*		Calibrated CPUE
	Longline fishery	Trotline fishery (calibrated to longline)	Longline index incorporating trotline data
1997	0.381		0.390
1998	0.212		0.214
1999	0.180		0.183
2000	0.250		0.251
2001	0.074		0.075
2002	0.132		0.134
2003	0.039		0.040
2004	0.133		0.135
2005	0.104		0.106
2006	0.096		0.097
2007	0.116		0.116
2008	0.150	0.081	0.137
2009	0.107	0.118	0.116
2010	0.123	0.171	0.141
2011	0.062	0.152	0.144
2012	0.069	0.099	0.081

Table 3. Exponentiated "gear" factor estimates from a GLMM that combined Spanish longline and trotline CPUE data, from the paired longline-trotline research data and from the calibration analysis, together with 95% CI's shown in brackets.

	From GLMM with all data	From paired longline-trotline research data	Estimated from calibration analysis
Gear			
Longline	1	1	1
Trotline	8.232 (7.219; 9.387)	6.437 (5.250; 7.892)	6.635 (5.513; 7.985)

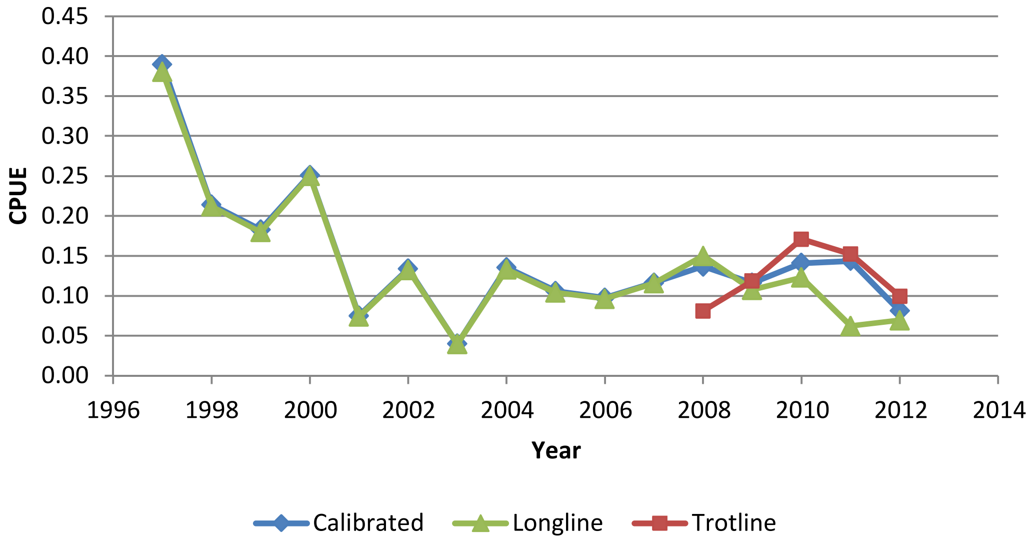

Figure 1. Calibrated longline CPUE trends as well as GLMM-standardised CPUE trends for the Spanish longline and trotline (calibrated to longline) toothfish fisheries for the Prince Edward Islands EEZ.

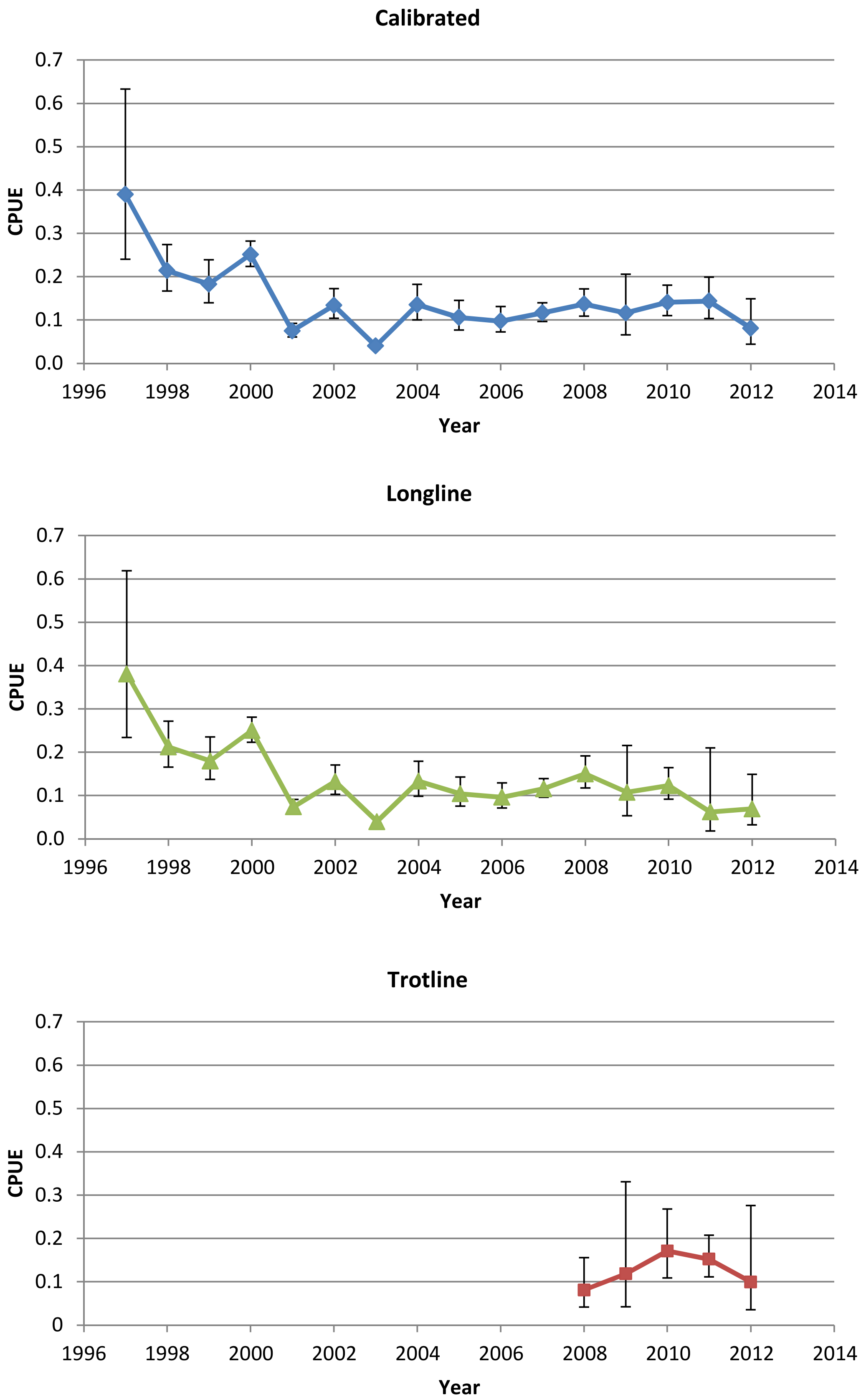

Figure 2. CPUE series of Figure 1 plotted individually with 95% CIs shown.